STAT 385: Methods for Data Analysis and System Optimization

Date: Spring 2015
Time: TTh 2:30 pm – 3:45 pm; Lab 6-8pm (Tues 20428 or Wed 21778)
Location: Duncan Hall, DCH 1064; Lab held in Dell Butcher Hall, DBH 214

Instructors

Prof. John A. Dobelman
dobelman@stat.rice.edu
Duncan Hall 2100, 713 348 5681
Office Hours: By appointment.

Teaching Assistant (TA)

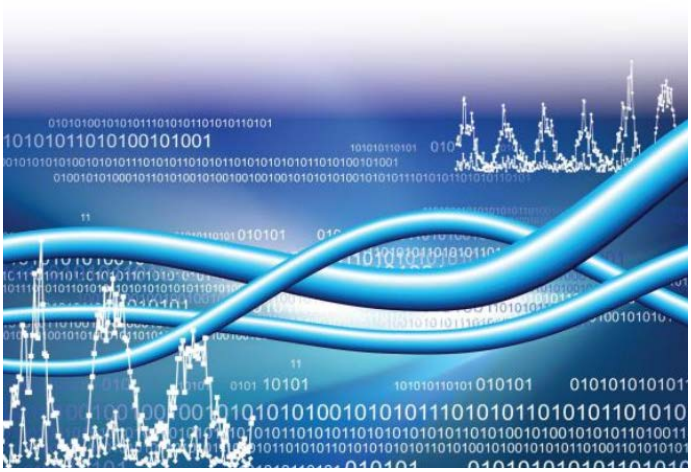
Steven M. Wilson
two@rice.edu
Duncan Hall 2090, 405-315-6165
Office Hours: By appointment

Course Text

Albright, Winston & Zappe (2011), *Data Analysis and Decision Making*, 4th Edition; ISBN-10: 0538476125 ISBN-13: 9780538476126
***(NOT** the 5th edition)*

Course Website

www.stat.rice.edu/~dobelman (see courses)
We will also be using OWL-Space



Course Description

The three general topic areas covered in this methodology oriented course are statistical methods including regression, sampling, and experimental design; simulation based methods in statistics, queuing and inventory problems; and an introduction to optimization methods. Excel will serve as the basic computing software

Every.Sentence.Here.Is.Important™

Sponsored Message Regarding Student Responsibility:

The Committee on Examinations and Standing has asked that we reiterate the responsibilities of the student to comply with deadlines affecting their status and standing. Essentially this means that you need to be aware of the deadlines for which you alone are responsible (i.e., not your advisor, etc.). This data is maintained on http://registrar.rice.edu/calendars/.

Textbook

Students are expected to have purchased their required texts during the first week of class. If the text is available at the "Rice" bookstore, and the student elects to purchase the book elsewhere, and this elsewhere-obtained book is not in hand, THE STUDENT IS NONETHELESS RESPONSIBLE for all assignments and readings. Some provision is made for down-edition situations, but advance coordination with the instructor/TA must be made. It is the student's responsibility to nonetheless comply with all deadlines if they choose to use a down-edition text. Additionally, Rice University is generally not able to provide copies of these texts for your use.

Grading

Grading for this course will consist of Labs and homework exercises/cases from the text (20%), concept quizzes (55%), a Final Team Project (20%), and Class Participation (5%)/Team Evaluation (5%, consisting of self-, peer- and instructor evaluations). If Team Projects are waived, the grade breakdown will be 20% Homework/Assignments, 75% Concept Quizzes, and 5% Class Participation. The final project is to be completed as a team project. It is expected that about 40% of the class will receive a grade equal to or higher than an "A-," with about a proportion of 7% receiving an "A+."

Assignment Submission and Late policy

All assignments must be submitted in class on the date due as hard copy, with a soft copy submitted on OWL-Space as well. Both hard and soft copies of the assignments are required in order to receive credit for the assignment, neither can be late. There are NO EXCEPTIONS. There will be a 20% penalty for HW turned in by next day; no credit for later than this. The instructors/TA will not be able to print out e-mailed or online-only assignment submissions. Please see the detailed submission standards for complete details.

NOTE: It is expected and encouraged that the student will use whatever software best helps them complete the assignments. Very little work will required hand-calculation of table lookup, although the student should know how to do this in case of shipwreck.

Laboratory

Occasionally, dedicated Lab projects will be assigned. Weekly lab time is set aside for you to work on and receive help on homework and team projects. Attendance is optional, but if you miss the lab problem-solving session, the TA will not be able provide the time to you.

Quizzes

There will be closed book quizzes given during the class period. These quizzes will cover concepts discussed in class. They will be announced in class and on Owl-space.

Final Team Project

A final project may be assigned which will be completed in groups/teams. Sometimes a quantitative case is provided, other times the project may be based on the various case studies included in your textbook. In the latter case, you and your team will choose a case to solve. You may pick your own team members, but the size of each team will be limited to an appropriate number (no more than three).

Extra Credit

From time to time, extra credit opportunities will be provided. These will be assigned in class, and generally NOT posted on Owl-Space.

Professional Standards

All assignments, with exception of handwritten math proofs and things of that nature, must be prepared nicely (like this syllabus, más o menos.) This means you need to organize it neatly with appropriate layout, graphics, citations, etc. Unnecessary printouts of numbers are not acceptable, nor are meaningless digits of precision, etc. The font should be large enough for management to easily read (no less than 10-point font, preferably 12).

For group projects you should include a cover page indicating the class, assignment, date submitted, and name of student or group members and group number. In the real world, if your management/customers or clients cannot read the report, or if it is filled with superfluous information, they will just return it to you with a request to resubmit. If you are unlucky, they will just pass you over for someone who does these things better.

Originality

Although some assignments are not "pledged", keep in mind that plagiarism is a serious problem, and is especially problematic for the student or researcher. Plagiarism will be treated, for pledged assignments, in accordance with the honor code provisions; and for non-pledged assignments, your paper will be returned for a rewrite with the automatic deduction of one letter grade. By plagiarism we mean "quoting, paraphrasing, or otherwise using another's words or ideas as one's own without properly crediting the source." Unfortunately, with the prevalence of internet sources, it is sometimes easy to inadvertently commit plagiarism. You should consult http://riceowl.rice.edu/guidance.cfm?doc_id=11767 or Academic-Fraud-and-the-Honor-System.pdf for guidance on how to avoid this problem.

Attendance

Students are expected to attend class. Much material is presented in class which might not otherwise be in texts, notes, etc. Attendance will be reflected in the Class Participation portion of the course grade. Although we plan on keeping the course website up to date, if a student misses a class, then he/she is responsible for keeping up with the course material and finding out if any exams, quizzes, or homeworks have been assigned or scheduled. Similarly, important due date changes might sometimes be made in class to your benefit which might not be immediately posted on Owl-Space.

Laptops and Other Portable Electronic Devices (PED)

Unless so requested by the instructors, please do not use these devices during class. These include laptops, earbuds, gaming devices, mp3/music/media players, cell/smart-phones, PDA's, Kindle/e-book readers, tablet computers, Apple i* devices, multi-purpose wrist communicators, cameras, GPS/GIS devices, etc. Hearing aids are acceptable to use in accordance with the last item (disabilities) and procedures herein. From time to time the instructors may ask a student to look something up, but in most cases this is not necessary. If the student requests, such devices may be permitted if the purpose is clearly articulated in advance. The student will be asked to put away their PED's if they are taken out in class; upon the second request, they will be collected by the instructor.

Software

The software for the course is primarily Microsoft Excel and the Palisades DecisionTools Suite Industrial, Textbook Edition. It is available for download (both for 3eR and 4th editions) at http://www.palisade.com/bookdownloads/albrightwinstonzappe/ Please note that these two versions are for Office 2010/2007 and later.

It is impossible to perform statistical/quantitative data analysis today without some sort of computer software, and it is expected that the student will become proficient with one or more statistical software packages. The most widely used data analysis software in the real world today is Microsoft Excel, and its capabilities are impressive. However, more specialized software is sometimes needed, such as Matlab, R, S-Plus, SAS, Resampling Stats, SPSS, Stata, StatTools, @Risk, Maple, Mathematica, C-Plex, SQL, etc. Of these, R has become the choice of many

because of its ease of use and low cost (FREE!). Download information for some of these packages is available on the course website(s).

One can also program most statistical procedures in a "high-level" programming language such as Java, C++, FORTRAN, VB, etc., along with specialized add-in routine libraries, but these require a lot of work to code and debug. Additionally, you will find that most corporate employers will not be paying for the nice software that is available for you here on campus. Consequently, to enhance your future value to your future employer, we suggest that you become proficient in Excel, R and SQL/SAS.

Course Content

We plan to cover most of this material. Note that Parts I-III have already been covered to some good extent in the prerequisites for this course, i.e. STAT 280 or STAT 305 or STAT 310 or ECON 382. If you do not remember these concepts, please review the material prior to coming to class since we cannot re-cover all this material in class. Topics/chapters include

Part 0.1: INTRODUCTION
1. Introduction. An Overview of the Book. The Methods. The Software. A Sampling of Examples. Modeling and Models.

Part I: GETTING, DESCRIBING, AND SUMMARIZING DATA.
2. Describing the Distribution of a Single Variable.
Basic Concepts. Descriptive Measures for Categorical Variables. Descriptive Measures for Numerical Variables. Time Series Data. Outliers and Missing Values. Excel Tables for Filtering, Sorting, and Summarizing.

3. Finding Relationships Among Variables.
Introduction. Relationships Among Categorical Variables. Relationships Among Categorical Variables and a Numerical Variable. Relationships Among Numerical Variables. Pivot Tables. An Extended Example.

Part II: PROBABILITY, UNCERTAINTY, AND DECISION MAKING.
4. Probability and Probability Distributions.
Introduction. Probability Essentials. Distribution of a Single Random Variable. An Introduction to Simulation. Distribution of Two Random Variables: Scenario Approach. Distribution of Two Random Variables: Joint Probability Approach. Independent Random Variables. Weighted Sums of Random Variables.

5. Normal, Binomial, Poisson, and Exponential Distributions.
Introduction. The Normal Distribution. Applications of the Normal Distribution. The Binomial Distribution. Applications of the Binomial Distribution. The Poisson and Exponential Distributions. Fitting a Probability Distribution to Data: BestFit.

6. Decision Making Under Uncertainty.
Introduction. Elements of a Decision Analysis. The PrecisionTree Add-In. Bayes' Rule. Multistage Decision Problems. Incorporating Attitudes Toward Risk.

Part III: STATISTICAL INFERENCE.
7. Sampling and Sampling Distributions.
Introduction. Sampling Terminology. Methods for Selecting Random Samples. An Introduction to Estimation.

8. Confidence Interval Estimation.
Introduction. Sampling Distributions. Confidence Interval for a Mean. Confidence Interval for a Total. Confidence Interval for a Proportion. Confidence Interval for a Standard Deviation. Confidence Interval for the Difference Between Means. Confidence Interval for the Difference Between Proportions Controlling Confidence Interval Length.

9. Hypothesis Testing.
Introduction. Concepts in Hypothesis Testing. Hypothesis Tests for a Population Mean. Hypothesis Tests for Other Parameters. Tests for Normality. Chi-Square Test for Independence. One-Way ANOVA.

Part IV: REGRESSION, FORECASTING, AND TIME SERIES.
10. Regression Analysis: Estimating Relationships.
Introduction. Scatterplots: Graphing Relationships. Correlations: Indicators of Linear Relationships Simple Linear Regression. Multiple Regression. Modeling Possibilities. Validation of the Fit.

11. Regression Analysis: Statistical Inference Introduction. The Statistical Model. Inferences About the Regression Coefficients. Multicollinearity. Include/Exclude Decisions. Stepwise Regression.The Partial F Test. Outliers. Violations of Regression Assumptions. Prediction.

12. Time Series Analysis and Forecasting.
Introduction. Forecasting Methods: An Overview. Testing for Randomness. Regression-Based Trend Models. The Random Walk Model. Autoregression Models. Moving Averages. Exponential Smoothing. Seasonal Models. Winters' Exponential Smoothing Model.

Part V: DECISION MODELING.
13. Introduction to Optimization Modeling.
Introduction. Introduction to Optimization. A Two-Variable Model. Sensitivity Analysis Properties of Linear Models. Infeasibility and Unboundedness. A Product Mix Model. A Multiperiod Production Model. A Comparison of Algebraic and Spreadsheet Models. A Decision Support System.

14. Optimization Modeling: Applications.
Introduction. Workforce Scheduling Models. Blending Models. Logistics Models. Aggregate Planning Models. Financial Models. Integer Programming Models. Nonlinear Models.

15.Introduction to Simulation Modeling.
Introduction. Real Applications of Simulation. Probability Distributions for Input Variables. Simulation with Built-In Excel Tools. Introduction to @RISK. The Effects of Input Distributions on Results.

16.Simulation Models.
Introduction. Operations Models. Financial Models. Marketing Models. Simulating Games of Chance.

Part 0.2: DECISION MODELING (CH 17).
17 Getting the Right Data.
Introduction. Sources of Data. Excel Tables for Filtering, Sorting, and Summarizing. Complex Queries with the Advanced Filter. Importing External Data from Access. Creating Pivot Tables from External Data. Web Queries. Other Data Sources on the Web. Cleansing the Data.

Rice Honor Code

Before enrolling in this course, you must understand and agree to abide by the Honor System in place at Rice University which protects the academic integrity of all coursework. All students (including graduate students) at Rice are bound by the Rice Honor Code. New students should familiarize themselves with the Honor Code before starting classes. The Honor Code is a unique feature at Rice, one that is valued highly, and is of profound importance. Honor Code violations are very serious, and can lead to dismissal from the University. See http://honor.rice.edu

Examinations, if held, are conducted under pledged conditions. Proper recitation and use of the Honor Pledge on examinations will normally be worth 5% of the exam grade.

Homework and group assignments/projects may be worked on with other class members but each student must submit their own work for credit. Homeworks should be submitted by each person, but you should indicate with whom you worked when applicable. No direct copying is allowed. Group projects and assignments may be submitted by the group.

Disabilities

Any student with a documented disability needing academic adjustments or accommodations is requested to speak with me during the first two weeks of class. All discussions will remain confidential. Students with disabilities should also contact Disability Support Services in the Ley Student Center. Further information is available at http://dss.rice.edu/.

Changes to Syllabus

Changes to this syllabus may be published from time to time, with notice and explanation given in class and via Owl-Space (if used).

Rev: 1/21/2015